Yatsko V.A.

Katanov State University of Khakasia
egalitt@gmail.com

Bayes theorem and a methodology for prediction of US presidential elections results

Abstract. *The paper develops a methodology for prediction of US presidential elections results. The methodology involves the following procedures. 1) Select time horizon for statistical analysis. 2) Select parameters for statistical analysis. 3) Study configuration of parameters during the election year for which the forecast is made and extrapolate this configuration to previous years. 4) Apply Bayes theorem using statistical data about previous elections to calculate probabilities for the GOP and Democrats. 5) If some or all parameters are not known make a projection for the given year basing on results of previous campaigns and developing optimistic, pessimistic and intermediate scenarios. Basing on this methodology the paper first makes prediction for 2016 US presidential election taking its results as a reference data and then formulates the projection for 2020 basing on optimistic and pessimistic scenarios for Republicans and Democrats.*

Keywords: US presidential elections, prediction methodology, Bayes theorem, statistical parameters

1. Introduction

Recent US presidential election aroused widespread engagement across the world, with many pollsters, politicians, and sorcerers making predictions about its outcome, mostly in Clinton's favor. *FiveThirtyEight*, a website that focuses on opinion poll analysis, predicted Clinton victory with 71.4% of electoral votes to Trump's 28.6[1]. A week before the election date *Moody's Analytics* asserted that Clinton would pick up 332 Electoral College votes against 206 for Trump [2]. Just hours before the election *Reuters/Ipsos States of the Nation* project assured that Clinton had about a 90 percent chance of defeating Trump in the race for the White House[3]. Sam Wang of *Princeton Election Consortium* gave an interview stating that there was 99% chance of Clinton win[4]. A couple of days before the election professor Wang promised to eat a bug if Trump would win more than 240 electoral votes[5]. French sorcerer and clairvoyant Kaede Uber also prophesized Clinton win[6].

In contrast to almost unanimous conviction of pollsters and experts that Clinton would win the election "The Keys to the White House" system, developed by Vladimir Keilis-Borok and Allan Lichtman and (Keilis-Borok, 1981], predicted Trump's victory. The distinctive feature of this system is that it focuses on results achieved by incumbent

[1] https://projects.fivethirtyeight.com/2016-election-forecast/
[2] http://www.reuters.com/article/us-usa-election-research-moody-s-idUSKBN12W56J
[3] http://www.reuters.com/article/us-usa-election-poll-idUSKBN1322J1
[4] http://www.msnbc.com/the-last-word/watch/scientist-predicts-99-chance-of-clinton-win-801634371744
[5] http://election.princeton.edu/category/2016-election/page/3/
[6] http://www.todaynews24h.com/human-communications-of-the-prophetess-vanga-guess-hillary-clinton-wins-the-u-s-presidential-election

administration as well as on data about previous elections rather than on events of the election campaign. In his book (Lichtman, 2008) Lichtman provides a detailed description of thirteen factors that determine the outcome of presidential elections in the USA. The system has worked well correctly predicting outcomes of presidential elections since 1984. One of the system's faults noted by experts is that it forecasts only the winner of elections without giving any specifics about vote share of the winning party (Silver, 2011).

The aim of this paper is to show that, basing on the factors distinguished by Lichtman and Keilis-Borok, Bayes theorem can be successfully applied to get specific numeric data about outcomes of US presidential elections.

During the last election Trump won the Electoral College with 304 votes (*r1*) compared to 227 votes (*r2*) for Hillary Clinton (Schmidt, 2016) thus getting 25.33% advantage according to the formula

$$Vm = 100 - \left(\frac{r2}{r1} * 100\right) \cong 25.33\% \quad (1)$$

where *Vm* is victory margin.

More common practice in modern social science is to correlate candidates' results with the total number of votes (c=538). Applying formula (2) we get 14.31% of victory margin.

$$Vm = \frac{r1}{c} * 100 - \frac{r2}{c} * 100 \cong 14.31\% \quad (2)$$

Joyce Berg et al (Berg, 2008) suggest a normalized margin of victory calculated by formula

$$Vm = \frac{r1 - r2}{r1 + r2} \cong 14.5\% \quad (3)$$

The normalized margin of victory[7] can be also found by formula

$$Vm = \frac{r1 - \frac{c}{2}}{\frac{c}{2}} \cong 13\% \quad (4)$$

Apparently, Trump's margin of victory ranges between 13-25 percent that can be considered the reference data range. Can we obtain these figures if we apply Bayes theorem basing on Lichtman's concept? To answer this question I will 1) consider distinctive features of the theorem; 2) analyze Lichtman's key points; 3) collect data needed for application of a Bayes formula; 4) get a result for 2016 presidential election and compare it with the reference range; 5) predict results of 2020 election taking into account different possible scenarios.

2. Bayes theorem and Lichtman's Keys

2.1. Bayes theorem is considered one of the foundations of contemporary probability theory (Garrick, 2008, p. 245) that has been widely used to: 1) make predictions and forecasts; 2) evaluate accuracy of tests, techniques, methodologies (mostly in medicine); 3) classify objects (mostly text documents). Table 1 presents formulas commonly used to address these issues.

[7] https://en.wikipedia.org/wiki/List_of_United_States_presidential_elections_by_Electoral_College_margin

Table 1. Formulas representing Bayes theorem

Formula	No	Formula's name	Issue addressed by the formula
$P(A\|B) = \frac{P(A) * P(B\|A)}{P(B)}$	(5)	Uni-denomial	Forecasting and predicting
$P(A\|B) = \frac{P(A\&B)}{P(B)}$	(6)	Concise uni-denomial	
$P(A\|B) = \frac{P(A) * P(B\|A)}{P(A) * P(B\|A) + P(B\|\sim A) * P(\sim A)}$	(7)	Bi-denomial	Evaluating
$\mathrm{P}(\mathrm{A}\|B_1, B_2 \ldots B_n) = \frac{\mathrm{P}(\mathrm{A}) * \ \mathrm{P}(B_1\|\mathrm{A}) * \mathrm{P}(B_2\|\mathrm{A}) * \ldots \mathrm{P}(B_n\|\mathrm{A})}{\mathrm{P}(B_1) * \mathrm{P}(B_2) * \ldots \mathrm{P}(B_n)}$	(8)	Poly-denomial	Forecasting and predicting
$P(x\|w_j) = \prod_{i=1}^{m} P(x_i\|w_j)$ (Raschka, 2014)	(9)	Multinomial	Classifying

There being no unanimously accepted classification, I named the first four formulas taking into account number of multipliers in denominator. As this paper addresses the problem of prediction of election results I will use formula (8), which is an extension of uni-denomial formula (5) formed by adding multipliers to numerator and denominator. In this formula *A* is victory of the Republican Party or Democratic party, and B_1, B_2.. B_n are parameters distinguished on the basis of Lichtman's approach. While analyzing these parameters one should bear in mind the following principles that underlie application of Bayes theorem.

1. The theorem rests on knowledge of prior events and calculation of prior probabilities. If the USA were a newly formed state, and in 2016 first election took place the theorem couldn't be applied. Fortunately the country has a long history of presidential elections, the results of which are available from various Internet resources. The most valuable of them is *Historical Presidential Elections*[8] that lists results of all US presidential elections from 1789 up to 2016. I will rely mostly on this resource to get statistical data needed for theorem's application. It should be noted that Lichtman's concept fits well this principle as it involves analysis of prior elections.

2. The theorem yields valid results on condition that it is applied to independent events. In this paper independent events are those ones that do not occur in the same year. If we don't follow this principle we are very likely to get probabilities whose values exceeds one, which is unacceptable.

[8] http://www.270towin.com/historical-presidential-elections/

3. The theorem can be applied to statistical data that involve frequencies of events. This principle is particularly important for the present investigation since not all of Lichtman's keys are of statistical nature.

2.2. Lichtman's keys

Table 2 presents thirteen Lichtman's keys indicating whether they are statistical or non-statistical. According to his system if six or more keys are false, the party in power will lose. As there are 6 false keys in the table (marked by "0") Lichtman predicted Trump's victory.

No	**Key event**	**Explanation**	**Lichtman's estimation**	**Statistical**	**Non statistical**
1.	Party Mandate	After the midterm elections, the incumbent party holds more seats in the U.S. House of Representatives than it did after the previous midterm elections.	0	+	
2.	Contest	There is no serious contest for the incumbent-party nomination.	1		+
3	Incumbency	The incumbent-party candidate is the sitting president.	0	+	
4	Third party	There is no significant third party or independent campaign.	0		+
5	Short-term economy	The economy is not in recession during the election campaign.	1	+	
6	Long-term economy	Real per-capita economic growth during the incumbent president's term equals or exceeds mean growth during the previous two terms.	1	+	
7	Policy change	The incumbent administration effects major changes in national policy.	0		+
8	Social unrest	There is no sustained social unrest during the term.	1		+
9	Scandal	The incumbent administration is untainted by major scandal.	1		+
10	Foreign/military failure	The incumbent administration suffers no major failure in foreign or military affairs.	1		+
11	Foreign/military success	The incumbent administration achieves a major success in foreign or military affairs.	0		+
12	Incumbent charisma	The incumbent-party candidate is charismatic or a national hero.	0		+
13	Challenger charisma	The challenging-party candidate is not charismatic or a national hero.	1		+

Statistical keys (or parameters) are those ones that have a distinct numerical basis. Real per capita growth can be easily estimated in specific figures in contrast to charisma of person. The same goes to phrases with vague quantifiers "serious contest", "major failure", "major success", "major changes", significant third party", and "sustained unrest". These parameters are non-statistical as they cannot be measured or calculated in specific figures. Moreover, they admit of different subjective interpretations. Lichtman is sure that Obama administration suffered no major failure in foreign or military affairs (key No 10). Kuperman, a *Foreign affairs* expert, considers Obama's intervention to Lybia a debacle and major failure (Kuperman, 2015). Obama himself admitted the intervention was his worst mistake[9]. The possibility of controversial interpretations excludes such keys from statistical analysis.

So, there are 4 parameters that comply with statistical analysis requirements. 1. Party Mandate. 2. Incumbency. 3. Short term economy. 4. Long term economy.

I will consider the first three of them. Analysis of the last parameter is complicated by the fact that economists seem to lack any unanimously accepted methodology to calculate real per capita GDP growth. Two sources, *United States GDP*[10] and *US Real GDP Per Capita by Year*[11] use the same methodology based on 2009 chained dollars but give different data with not a single identical figure. For example, *United States GDP* estimates real per capita growth in 2008 at $48401, while *US Real GDP Per Capita by Year* gives for the same year real per capita growth of $49206.50. *Social Democracy for the 21st Century*[12] and Kimberly Amadeo (Amadeo, 2017) offer some other methodologies that produce different results.

Long term economy parameter in Lichtman's interpretation (if I've got it properly) can be calculated by the formula

$$G = (\mathrm{C} - Y_1) - ((Y_1 - Y_8)/2) \tag{10}$$

According to (Amadeo, 2017) in 2012, first year of Obama's last term (Y_1), real GDP per capita was $49,039, and in 2016 ($C$) it was $57,300. Growth during the period 2004-2012 is estimated at $1239 ($47,800 in 2004 (Y_8)). If we use these data we get growth

$$G = (57300 - 49039) - ((49039 - 47800)/2) = 7641.5 \tag{11}$$

But data from *United States GDP* yield and entirely different result: G=*-801*. Apparently such contradictory approaches cannot be used to make reliable forecasts.

3. Prediction for 2016

In statistical analysis one of the problems is to correctly choose the time horizon, for which the data are collected. I will take the period 1952-2016 basing on the following

[9] https://www.theatlantic.com/international/archive/2016/204/obamas-worst-mistake-libya/478461/
[10] http://www.deptofnumbers.com/gdp/us/ (last accessed July 9, 2017)
[11] http://www.multpl.com/us-real-gdp-per-capita/table/by-year
[12] http://socialdemocracy21stcentury.blogspot.ru/2012/09/us-real-per-capita-gdp-from-18702001.html

assumptions. Firstly, Eisenhower's win in 1952 ended a long 20-year Democratic Presidents rule. Actually, such a long term was unprecedented for American politics and could be accounted for by the two biggest crises in US 20th century history: The Great Depression and World War II. Eisenhower's coming to power marked reversion to normal political process characterized by presidency alternation between Republicans and Democrats. Secondly, the beginning of 1950s was distinguished by US-Russia Cold War tensions and since that time US-Russia relations have become an important issue of all presidential campaigns, the 2016 campaign being in no way an exception. Thirdly, it was in the 1950s that modern technologies demonstrated their impact on political processes in the USA. Televisions became a standard household appliance around the country allowing politicians to make direct connections with voters during election campaigns (Green, 2017). Nowadays, as 2016 campaign clearly demonstrated, such direct connections realized on the basis of modern computer technologies may play a decisive role for elections outcome.

In (Mayhew, 2008) the first parameter is referred to as incumbency advantage that implies advantage of the party that controls either the House or the House and the Senate. Consequently, I will take the liberty to reformulate the definition on this parameter: party mandate involves control by the given party over the House or the House and the Senate after a midterm election.

Tables 3 and 4 provide statistical data about midterm and presidential elections in the USA

Table 3. Statistics for midterm elections[13]

No	Year	B1R		B1D	
		+	-	+	-
1	1950		-	+	
2	1954		-	+	
3	1958		-	+	
4	1962		-	+	
5	1966		-	+	
6	1970		-	+	
7	1974		-	+	
8	1978		-	+	
9	1982		-	+	
10	1986		-	+	
11	1990		-	+	
12	1994	+			-
13	1998	+			-
14	2002	+			-
15	2006		-	+	
16	2010		-	+	
Total		3	13	13	3

[13] https://en.wikipedia.org/wiki/Republican_Party_(United_States)

Table 4. Statistics for presidential elections[14]

No	Year	B2R		B2D		B2SR		B2SD	
		+	-	+	-	+	-	+	-
1	1952	+			-		-		-
2	1956	+			-	+			-
3	1960		-	+			-		-
4	1964		-	+			-		-
5	1968	+			-		-		-
6	1972	+			-	+			-
7	1976		-	+			-		-
8	1980	+			-		-	+	
9	1984	+			-	+			-
10	1988	+			-		-		-
11	1992		-	+		+			-
12	1996		-	+			-	+	
13	2000	+			-		-		-
14	2004	+			-	+			-
15	2008		-	+			-		-
16	2012		-	+			-	+	
Total		9	7	7	9	5	11	3	13

In these tables:
B1R – Republican Party mandate;
B1D – Democratic Party mandate;
B2R - Republican Party wins election;
B2D - Democratic Party wins election;
B2SR – sitting Republican president runs for reelection;
B2SD - sitting Democratic president runs for reelection.

Table 4 shows that Republicans won 9 out of 16 presidential elections while Democrats won 7 times. Note that results of midterm elections in 2014 won by Republicans are not counted as they are correlated with 2016 presidential election, for which we make the forecast, and I included in Table 3 results of 1950 midterm elections instead. Nevertheless we must bear them in mind because they set up a configuration for statistical data collection. To correctly estimate both parties' chances during 2016 presidential election one must calculate how many times Republicans won having majority in the House and how many times Democrats won being in minority in the House. Republicans controlled the House in 1994, 1998, and 2002, winning presidential elections in 2000 and 2004, which gives 2/3=0.666667≅67%. Democrats had minority in the same years winning in 1996, which yields 1/3=0.33333≅33%.

Uni-denomial formulas[15]:

[14] http://www.270towin.com/historical-presidential-elections/
[15] Figures rounded automatically in MS Excel, general cells format

$$\mathrm{P}(\mathrm{AR}|B_1R) = \frac{\mathrm{P(AR)} * \mathrm{P}(B_1R|\mathrm{AR})}{\mathrm{P}(B_1R)} = \frac{0.5625*0.22222}{0.1875} = 0.666667 \cong 66.67\% \quad (12)$$

$$\mathrm{P}(\mathrm{AD}|B_1D) = \frac{\mathrm{P(AD)} * \mathrm{P}(B_1D|\mathrm{AD})}{\mathrm{P}(B_1D)} = \frac{0.4375*0.142857}{0.1875} = 0.3333 \cong 33.33\% \quad (13)$$

Republicans perform much better when their sitting presidents run for reelection. They won 4 times out of 5 times (Bush was defeated by Clinton in 1992). Democratic candidates-presidents won 2 times out of 3 (Carter lost to Reagan in 1980). But in 2016 election, for which we make our forecast, none of the candidates was incumbent president and we must consider a "non-incumbency" parameter, i.e. elections when Republican and Democratic candidates were not sitting presidents. Republicans won 5 out of 11 such elections and Democrats - 5 of 13. In the formulas below B_2 denotes probabilities for candidates who weren't incumbent presidents.

$$\mathrm{P}(\mathrm{AR}|B_2\mathrm{R}) = \frac{\mathrm{P(AR)} * \mathrm{P}(B_2R|\mathrm{AR})}{\mathrm{P}(B_2\mathrm{R})} = \frac{0.5625*0.55555556}{0.69} = 0.454545 \cong 45.45\% \quad (14)$$

$$\mathrm{P}(\mathrm{AD}|B_2\mathrm{D}) = \frac{\mathrm{P(AD)} * \mathrm{P}(B_2\mathrm{D}|\mathrm{AD})}{\mathrm{P}(B_2\mathrm{D})} = \frac{0.44*0.714299}{0.8125} = 0.38462 \cong 38.46\% \quad (15)$$

The results don't make up 100% sum because some data go to elections with candidates-presidents. Consequently, we will have to calculate probabilities for two parties separately rather than first obtaining the result for one party and then find the result for the other party as 100%-first party's result. The calculations can be verified by concise formulas.

$$\mathrm{P}(\mathrm{AR}|B_2\mathrm{R}) = \frac{\mathrm{P}(\mathrm{AR}\&B_2\mathrm{R})}{\mathrm{P(niR)}} = \frac{5}{11} = 0.454545 \cong 45.45\% \quad (16)$$

$$\mathrm{P}(\mathrm{AD}|B_2\mathrm{D}) = \frac{\mathrm{P}(\mathrm{AD}\&B_2\mathrm{D})}{\mathrm{P}(B_2\mathrm{D})} = \frac{5}{13} = 0.38462 \cong 38.46\% \quad (17)$$

Let's proceed to the last parameter. Economic situation was favorable for Democrats, the economy was not in recession during the 2016 election campaign. The task is to find statistics about chances of Republican victory under such circumstances. Recession is unanimously interpreted as two consecutive quarters of negative economic growth as measured by a country's gross domestic product[16]. Table 5 presents data about this parameter.

Table 5. Data about recession[17]

No	Year	Election won by		Recession	
1	1952	R		-	
2	1956	R		-	
3	1960		D		+
4	1964		D	-	

[16] http://www.investopedia.com/terms/r/recession.asp
[17] https://www.thebalance.com/us-gdp-by-year-3305543

5	1968	R		-	
6	1972	R		-	
7	1976		D	-	
8	1980	R			+
9	1984	R		-	
10	1988	R		-	
11	1992		D	-	
12	1996		D	-	
13	2000	R		-	
14	2004	R		-	
15	2008		D		+
16	2012		D	-	
Total		9	7	13	3

As the table shows recessions occurred 3 times during presidential election campaigns, Republicans winning once during recession 1980 and 8 times when there was no recession, i.e. 8/13=0.61538.

Unidenomial formula:

$$\mathrm{P}(\mathrm{AR}|B_3R) = \frac{\mathrm{P}(\mathrm{AR})*\mathrm{P}(B_3R|\mathrm{AR})}{\mathrm{P}(B_3R)} = \frac{0.5625*0.88889}{0.8125} = 0.61538 \cong 61.54\% \quad (18)$$

Democrats won 5 elections when there was no recession, i.e. 5/13=0.38462.

$$\mathrm{P}(\mathrm{AD}|B_3D) = \frac{\mathrm{P}(\mathrm{AD})*\mathrm{P}(B_3D|\mathrm{AD})}{\mathrm{P}(B_3D)} = \frac{0.4375*0.71429}{0.8125} = 0.38462 \cong 38.46\% \quad (19)$$

Finally, let's calculate probabilities for the GOP and Democrats using the three parameters.

$$\mathrm{P}(\mathrm{AR}|B_1R, B_2R, B_3R) = \frac{\mathrm{P}(\mathrm{AR})*\mathrm{P}(B_1R|\mathrm{AR})*\mathrm{P}(B_2R|\mathrm{AR})*\mathrm{P}(B_3R|\mathrm{AR})}{\mathrm{P}(B_1R)*\mathrm{P}(B_2R)*\mathrm{P}(B_3R)} = 0.5893695 \cong 58.94\% \quad (20)$$

$$\mathrm{P}(\mathrm{AD}|B_1D, B_2D, B_3D) = \frac{\mathrm{P}(\mathrm{AD})*\mathrm{P}(B_1D|\mathrm{AD})*\mathrm{P}(B_2D|\mathrm{AD})*\mathrm{P}(B_3D|\mathrm{AD})}{\mathrm{P}(B_1D)*\mathrm{P}(B_2D)*\mathrm{P}(B_3D)} = 0.25762 \cong 25.76\% \quad (21)$$

These results prove that in 2016 election campaign Republicans' advantage over Democrats can be estimated at about 33.18%. This figure doesn't fit well the reference range of victory margins defined earlier, being about 10% bigger. I will take this error rate into account to apply a correction to the 2020 forecast. The correction is calculated the median of the numerical series that changes from 13 to 25, range of Trump's victory margins. Apparently, the median for this series equals 19.

4. Forecast for 2020 election

To make a projection for next US presidential election one should proceed from different scenarios. The optimistic scenario for the GOP involves the following parameters.

1. Republicans control the House after 2018 midterm election (B_1).
2. Trump runs for reelection (B_2).
3. There is no recession in 2020 (B_3).

Accordingly the GOP loosing majority in the House, Trump's not running for reelection and recession during election year are parameters of pessimistic scenario.

Basing on the data found earlier chances of Republicans calculated according optimistic scenario are the following.

$$\mathrm{P}(\mathrm{AR}|B_1R) = \frac{\mathrm{P}(\mathrm{AR})*\mathrm{P}(B_1R|\mathrm{AR})}{\mathrm{P}(B_1R)} = \frac{0.58823589*0.3}{0.2941176} = 0.6 = 60\% \quad (22)$$

$$\mathrm{P}(\mathrm{AR}|B_2\mathrm{R}) = \frac{\mathrm{P}(\mathrm{AR})*\mathrm{P}(B_2|\mathrm{AR})}{\mathrm{P}(B_2\mathrm{R})} = \frac{0.58823589*0.4}{0.299412} = 0.8 = 80\% \quad (23)$$

$$\mathrm{P}(\mathrm{AR}|B_3R) = \frac{\mathrm{P}(\mathrm{AR})*\mathrm{P}(B_3R|\mathrm{AR})}{\mathrm{P}(B_3R)} = \frac{0.58823589*0.9}{0.8235294} = 0.642857 \cong 64.29\% \quad (24)$$

$$\mathrm{P}(\mathrm{AR}|B_1R, B_2R, B_3R) = \frac{\mathrm{P}(\mathrm{AR})*\mathrm{P}(B_1R|\mathrm{AR})*\mathrm{P}(B_2R|\mathrm{AR})*\mathrm{P}(B_3R|\mathrm{AR})}{\mathrm{P}(B_1R)*\mathrm{P}(B_2R)*\mathrm{P}(B_3R)} = 0.89177143 \cong 89.18\% \quad (25)$$

Note that the priors have changed as 2016 election is counted. The total number of elections is 17 and Republicans won 10 times.

Pessimistic scenario.

$$\mathrm{P}(\mathrm{AR}|B_1R) = \frac{\mathrm{P}(\mathrm{AR})*\mathrm{P}(B_1R|\mathrm{AR})}{\mathrm{P}(B_1R)} = \frac{0.58823589*0.8}{0.705882} = 0.66667 \cong 66.67\% \quad (26)$$

$$\mathrm{P}(\mathrm{AR}|B_2\mathrm{R}) = \frac{\mathrm{P}(\mathrm{AR})*\mathrm{P}(B_2|\mathrm{AR})}{\mathrm{P}(B_2\mathrm{R})} = \frac{0.58823589*0.6}{0.705882} = 0.5 = 50\% \quad (27)$$

$$\mathrm{P}(\mathrm{AR}|B_3R) = \frac{\mathrm{P}(\mathrm{AR})*\mathrm{P}(B_3R|\mathrm{AR})}{\mathrm{P}(B_3R)} = \frac{0.58823589*0.1}{0.176471} = 0.33333 \cong 33.33 \quad (28)$$

$$\mathrm{P}(\mathrm{AR}|B_1R, B_2R, B_3R) = \frac{\mathrm{P}(\mathrm{AR})*\mathrm{P}(B_1R|\mathrm{AR})*\mathrm{P}(B_2R|\mathrm{AR})*\mathrm{P}(B_3R|\mathrm{AR})}{\mathrm{P}(B_1R)*\mathrm{P}(B_2R)*\mathrm{P}(B_3R)} = 0.3211111 \cong 32.11\% \quad (29)$$

The scenarios for Democrats are dependent on those ones for Republicans. If Republican's pessimistic scenario is realized optimistic scenario for Democrats involves:

1. Democrats control the House after 2018 midterm election (B_1).
2. Trump doesn't run for reelection. Neither of the candidates is a sitting president (B_2).
3. There is recession in 2020 (B_3).

$$\mathrm{P}(\mathrm{AD}|B_1D) = \frac{\mathrm{P}(\mathrm{AD})*\mathrm{P}(B_1D|\mathrm{AD})}{\mathrm{P}(B_1D)} = \frac{0.41*0.8571429}{0.7647059} = 0.4615385 \cong 46.15\% \quad (30)$$

$$\mathrm{P}(\mathrm{AD}|B_2\mathrm{D}) = \frac{\mathrm{P}(\mathrm{AD})*\mathrm{P}(B_2\mathrm{D}|\mathrm{AD})}{\mathrm{P}(B_2\mathrm{D})} = \frac{0.41*0.71428571}{0.82352941} = 0.3571429 \cong 35.71\% \quad (31)$$

$$\mathrm{P}(\mathrm{AD}|B_3D) = \frac{\mathrm{P}(\mathrm{AD})*\mathrm{P}(B_3D|\mathrm{AD})}{\mathrm{P}(B_3D)} = \frac{0.41*0.28571429}{0.1764706} = 0.6666667 \cong 66.67\% \quad (32)$$

$$\mathrm{P}(\mathrm{AD}|B_1D, B_2D, B_3D) = \frac{\mathrm{P}(\mathrm{AD})*\mathrm{P}(B_1D|\mathrm{AD})*\mathrm{P}(B_2D|\mathrm{AD})*\mathrm{P}(B_3D|\mathrm{AD})}{\mathrm{P}(B_1D)*\mathrm{P}(B_2D)*\mathrm{P}(B_3D)} = 0.64812738 \cong 64.81\% \quad (33)$$

If Republican's optimistic scenario is realized pessimistic scenario for Democrats involves:

1. Republicans control the House after 2018 midterm election (B_1).
2. Trump runs for reelection (B_2).
3. There is no recession in 2020 (B_3).

$$\mathrm{P}(\mathrm{AD}|B_1D) = \frac{\mathrm{P}(\mathrm{AD})*\mathrm{P}(B_1D|\mathrm{AD})}{\mathrm{P}(B_1D)} = \frac{0.41*0.71428571}{0.82352941} = 0.25 = 25\% \quad (34)$$

$$\mathrm{P}(\mathrm{AD}|B_2\mathrm{D}) = \frac{\mathrm{P}(\mathrm{AD})*\mathrm{P}(B_2\mathrm{D}|\mathrm{AD})}{\mathrm{P}(B_2\mathrm{D})} = \frac{0.41*0.1428571}{0.294118} = 0.2 = 20\% \quad (35)$$

$$\mathrm{P}(\mathrm{AD}|B_3D) = \frac{\mathrm{P}(\mathrm{AD})*\mathrm{P}(B_3D|\mathrm{AD})}{\mathrm{P}(B_3D)} = \frac{0.41*0.71428571}{0.76470588} = 0.3846154 \cong 38.46\% \quad (36)$$

$$\mathrm{P}(\mathrm{AD}|B_1D, B_2D, B_3D) = \frac{\mathrm{P}(\mathrm{AD})*\mathrm{P}(B_1D|\mathrm{AD})*\mathrm{P}(B_2D|\mathrm{AD})*\mathrm{P}(B_3D|\mathrm{AD})}{\mathrm{P}(B_1D)*\mathrm{P}(B_2D)*\mathrm{P}(B_3D)} = 0.11342229 \cong 11.34\% \quad (37)$$

So according to optimistic scenario for Republicans and pessimistic scenario for Democrats advantage of the former over the latter is *Vm* = 89.18% - 11.34% = 77.84%. Applying the correction we get *Vm* = 77.84% - 19 = 58.84%. Under pessimistic scenario for Republicans and optimistic scenario for Democrats the latter's advantage is *Vm* = 64.81% - 32.11% = 34.56. Of course other mixed scenarios are possible, such as: Trump runs for re-election but Republicans lose control of the House; Democrats control the House but there is recession, etc. No problem to calculate probabilities of their victories given the methodology describes in this paper.

5. Conclusions

Perhaps one of the most striking features of 2016 presidential election in the USA was bad mistakes made by various polling agencies and pollsters, all of which failed to correctly predict the election outcome. The failure is explained by two main factors: many of Trump backers did not respond to surveys, and many of them, the so-called "shy Trumpers" gave false information about their election preferences because support for Trump was considered socially undesirable (Mercer, 2016). Hence high subjectivity and error rates of polls.

This paper adopts an original statistical approach based on Bayes theorem and suggests the following methodology for making forecasts about US presidential elections results.

1. Select time horizon for statistical analysis. I suggest that this horizon should open at the beginning of fifties, time marked by a number of features described above that are characteristic of contemporary socio-political situation in the USA.

2. Select parameters for statistical analysis. Based on Lichtman's concept I selected three such parameters but this choice is in no way ultimate and is subject for further discussion. The parameters, as I see them, may be interpreted as binary oppositions and presented in terms of componential analysis (Kay, 1975) as ±party mandate, ±incumbency,

±recession. The first two parameters are distributed among parties and the third one is distributed across time. Interestingly, if only the first two parameters (party mandate and incumbency) are counted the result fits well the reference range, Trump's advantage estimated at 16%.

$$\mathrm{P}(\mathrm{AR}|B_1R, B_2R) = \frac{\mathrm{P}(\mathrm{AR})*\mathrm{P}(B_1R|\mathrm{AR})*\mathrm{P}(B_2R|\mathrm{AR})}{\mathrm{P}(B_1R)*\mathrm{P}(B_2R)} = 0.45584 \cong 45.58\% \quad (38)$$

$$\mathrm{P}(\mathrm{AD}|B_1D, B_2D) = \frac{\mathrm{P}(\mathrm{AD})*\mathrm{P}(B_1D|\mathrm{AD})*\mathrm{P}(B_2D|\mathrm{AD})}{\mathrm{P}(B_1D)*\mathrm{P}(B_2D)} = 0.293 \cong 29.3\% \quad (39)$$

There is obviously some interdependency between parameters and, for example, sitting president's refusal to run for reelection may be accounted for by bad economic situation including recession. But analysis of this interdependency and ranking of parameters according to their significance falls out of scope of this paper, being one of the important tasks of political science.

3. Study configuration of parameters during the election year for which the forecast is made and extrapolate this distribution to previous years and election campaigns. If, for example, in this year one of the candidates is sitting president the task is to find number of times when candidates from the given party ran for presidency and number of times they won. For the other party statistics about performance of candidates that were not incumbent presidents is found.

4. Apply Bayes theorem using statistical data about previous elections to calculate probabilities for the GOP and Democrats. I used poly-denomial formulas verifying it with the help of uni-denomial formulas.

5. If some or all parameters are not known it's possible to make a projection for the given year basing on results of previous campaigns and developing optimistic and pessimistic scenarios. The scenarios developed in this paper show that Republicans have better chances to win in 2020 mostly due to Trump's possible run for reelection. Being well aware of this fact the Democratic establishment wages a fierce struggle aimed at defamation of incumbent president and his impeachment.

Thus the statistical analysis helps to better understand current political situation in the USA

Acknowledgements

The author expresses gratitude to participants *Atlas Forum*[18] for valuable information about US elections.

This research is supported by a grant from Russian Foundation for Basic Research No 16-07-00014.

[18] http://uselectionatlas.org/FORUM/index.php

References

1. Amadeo, K. (2017) Real GDP per capita: How to calculate, data since 1946 https://www.thebalance.com/g00/real-gdp-per-capita-how-to-calculate-data-since-1946-3306028?i10c.referrer= (last accessed July 9, 2017)**.**

2. Berg, J.E. Nelson, F.D. and Rietz, T.A. (2008) Prediction market accuracy in the long run. *International Journal of Forecasting* 24 (2): 285–300. doi:10.1016/j.ijforecast.2008.03.007

3. Garrick, J. B. (2008) Quantifying and controlling catastrophic risks. Academic Press,

4. Green, J. (2017) How did television change the politics of the 1950s? http://classroom.synonym.com/did-television-change-politics-1950s-11086.html (last accessed July 9, 2017)

5. Kay, C., Samuels M.L. (1975) Componential analysis in semantics: It's validity and applications. Transactions of the Philological Society, 74 (1): 49-81 DOI: 10.1111/j.1467-968X.1975.tb01167.x

6. Keilis-Borok, V. I. and Lichtman, A. J. (1981). Pattern recognition applied to presidential elections in the United States, 1860–1980: The role of integral social, economic, and political traits. *Proceedings of the National Academy of Sciences*. 78 (11): 7230-7234. doi:10.1073/pnas.78.11.7230

7. Kuperman, A. J. (2015) Obama's Libya debacle. How a well-meaning intervention ended in failure. Foreign Affairs. March/April 2015. https://www.foreignaffairs.com/articles/libya/obamas-libya-debacle (last accessed July 9, 2017)

8. Lichtman, A. J. (2008) The keys to the White House. A surefire guide to predicting the next president. Lanham, MD: Rowman & Littlefield

9. Mayhew, D. R. (2008) Incumbency advantage in U.S. presidential elections: The historical record. *Political science quarterly*. 123 (2): 201-228. http://campuspress.yale.edu/davidmayhew/files/2013/05/mayhew_incumbencyadvantage_2008.pdf (last accessed July 9, 2017)

10. Mercer A., Deane C., and McGeeny, K. (2016) Why 2016 election polls missed their mark. *FactTank*, November 9, 2016. http://www.pewresearch.org/fact-tank/2016/11/09/why-2016-election-polls-missed-their-mark/ (last accessed July 9, 2017)

11.Raschka, S. (2014) Naive Bayes and text classification – introduction and theory. http://sebastianraschka.com/Articles/2014_naive_bayes_1.html (last accessed July 9, 2017)

12. Schmidt, K. and Andrews, W. (2016) A Historic number of electors defected, and most were supposed to vote for Clinton. *The New York Times*, December 19, 2016. https://www.nytimes.com/interactive/2016/12/19/us/elections/electoral-college-results.html

13. Silver, N. (2011) Despite keys Obama is no lock. *FiveThirtyEight*, August 31, 2016. https://fivethirtyeight.blogs.nytimes.com/2011/08/31/despite-keys-obama-is-no-lock/ (last accessed July 9, 2017)